Google Cloud Platform Ingénierie de données

Informatique

Cloud / DevOps

Référence : 4-GO-INDO
Durée : 4 jours
Présentiel ou en classe à distance

Tarif Inter : 750 € Prix HT jour / personne
Tarif Intra : 1400 € Prix HT jour / groupe

Mise à jour : 27/11/2023

Durée de validité : du 01/01/2026 au 31/12/2026

Objectifs

- Apprendre à concevoir et déployer des pipelines et des architectures pour le traitement des données
- Comprendre comment créer et déployer des workflows de machine learning
- Être capable d'interroger des ensembles de données
- Comprendre comment visualiser des résultats des requêtes et créer des rapports

Prérequis

Aucun prérequis obligatoire

Niveau recommandé

- Maîtriser les principes de base des langages de requête courants tels que SQL
- Avoir de l'expérience en modélisation, extraction, transformation et chargement des données
- Savoir développer des applications à l'aide d'un langage de programmation courant tel que Python
- Savoir utiliser le Machine Learning et/ou les statistiques

Public concerné

Développeurs expérimentés en charge des transformations du Big Data

Contenu pédagogique

Introduction à l'ingénierie des données

- Explorer le rôle d’un data engineer
- Analyser les défis d’ingénierie des données
- Introduction à BigQuery
- Data lakes et data warehouses
- Démo: requêtes fédérées avec BigQuery
- Bases de données transactionnelles vs data warehouses
- Démo: recherche de données personnelles dans votre jeu de données avec l’API DLP
- Travailler efficacement avec d’autres équipes de données
- Gérer l’accès aux données et gouvernance
- Construire des pipelines prêts pour la production
- Etude de cas d’un client GCP
- Lab : Analyse de données avec BigQuery

Construire un Data lake

- Introduction aux data lakes
- Stockage de données et options ETL sur GCP
- Construction d’un data lake à l’aide de Cloud Storage
- Démo : optimisation des coûts avec les classes et les fonctions cloud de Google Cloud Storage
- Sécurisation de Cloud Storage
- Stocker tous les types de données
- Démo : exécution de requêtes fédérées sur des fichiers Parquet et ORC dans BigQuery
- Cloud SQL en tant que data lake relationnel

Construire un Data Warehouse

- Le data warehouse moderne
- Introduction à BigQuery
- Démo : Requêter des TB + de données en quelques secondes
- Commencer à charger des données
- Démo: Interroger Cloud SQL à partir de BigQuery
- Lab : Chargement de données avec la console et la CLI
- Explorer les schémas
- Exploration des jeux de données publics BigQuery avec SQL à l'aide de Information_Schema
- Conception de schéma
- Démo : Exploration des jeux de données publics BigQuery avec SQL à l'aide de Information_Schema
- Champs imbriqués et répétés dans BigQuery
- Lab : tableaux et structures
- Optimiser avec le partitionnement et le clustering
- Démo : Tables partitionnées et groupées dans BigQuery
- Aperçu : Transformation de données par lots et en continu

Introduction à la construction de pipelines de données par lots EL, ELT, ETL

- Considérations de qualité
- Comment effectuer des opérations dans BigQuery
- Démo : ETL pour améliorer la qualité des données dans BigQuery
- Des lacunes
- ETL pour résoudre les problèmes de qualité des données

Exécution de Spark sur Cloud Dataproc

- L'écosystème Hadoop
- Exécution de Hadoop sur Cloud Dataproc GCS au lieu de HDFS
- Optimiser Dataproc
- Atelier : Exécution de jobs Apache Spark sur Cloud Dataproc

Traitement de données sans serveur avec Cloud dataflow

- Cloud Dataflow
- Pourquoi les clients apprécient-ils Dataflow ?
- Pipelines de flux de données
- Lab : Pipeline de flux de données simple (Python / Java)
- Lab : MapReduce dans un flux de données (Python / Java)
- Lab : Entrées latérales (Python / Java)
- Templates Dataflow
- Dataflow SQL

Gestion des pipelines de données avec Cloud Data fusion and Cloud composer

- Création visuelle de pipelines de données par lots avec Cloud Data Fusion: composants, présentation de l'interface utilisateur, construire un pipeline, exploration de données en utilisant Wrangler
- Lab : Construction et exécution d'un graphe de pipeline dans Cloud Data Fusion
- Orchestrer le travail entre les services GCP avec Cloud Composer - Apache Airflow
- Environment : DAG et opérateurs, planification du flux de travail
- Démo : Chargement de données déclenché par un événement avec Cloud Composer, Cloud Functions, Cloud Storage et BigQuery
- Lab : Introduction à Cloud Composer

Introduction au traitement de données en streaming

- Traitement des données en streaming

Serverless messaging avec Cloud Pub/Sub

- Cloud Pub/Sub

- Lab : Publier des données en continu dans Pub/Sub

Fonctionnalités streaming de Cloud Dataflow

- Fonctionnalités streaming de Cloud Dataflow
- Lab : Pipelines de données en continu

Fonctionnalités streaming à haut débit BIGQUERY ET BIGTABLE

- Fonctionnalités de streaming BigQuery
- Lab : Analyse en continu et tableaux de bord
- Cloud Bigtable
- Lab : Pipelines de données en continu vers Bigtable

Fonctionnalités avancées de BIGQUERY et performance

- Analytic Window Functions
- Utiliser des clauses With
- Fonctions SIG
- Démo: Cartographie des codes postaux à la croissance la plus rapide avec BigQuery GeoViz
- Considérations de performance
- Lab : Optimisation de vos requêtes BigQuery pour la performance
- Lab : Création de tables partitionnées par date dans BigQuery

Introduction à l'analytique et à l'IA

- Qu’est-ce que l’IA?
- De l’analyse de données ad hoc aux décisions basées sur les données
- Options pour modèles ML sur GCP

API de modèle ML prédéfinis pour les données non structurées

- Les données non structurées sont difficiles à utiliser
- API ML pour enrichir les données
- Lab : Utilisation de l’API en langage naturel pour classer le texte non structuré

Big Data Analytics avec les notebooks Cloud AI plateform

- Qu’est-ce qu’un notebook
- BigQuery Magic et liens avec Pandas
- Lab : BigQuery dans Jupyter Labs sur IA Platform

Pipeline de production ML avec Kubeflow

- Façons de faire du ML sur GCP
- Kubeflow AI Hub
- Lab : Utiliser des modèles d’IA sur Kubeflow

Création de modèles personnalisés avec SQL dans BIGQUERY ML

- BigQuery ML pour la construction de modèles rapides
- Démo : Entraîner un modèle avec BigQuery ML pour prédire les tarifs de taxi à New York
- Modèles pris en charge
- Lab : Prédire la durée d’une sortie à vélo avec un modèle de régression dans BigQuery ML
- Lab : Recommandations de film dans BigQuery ML

Création de modèles personnalisés avec Cloud AUTOML

- Pourquoi Auto ML?
- Auto ML Vision
- Auto ML NLP
- Auto ML Tables

Moyens pédagogiques

- Réflexion de groupe et apports théoriques du formateur.

- Travail d'échange avec les apprenants sous forme de réunion - discussion.
- Utilisation de cas concrets issus de l'expérience professionnelle.
- Validation des acquis par des questionnaires, des tests d'évaluation, des mises en situation et des jeux pédagogiques.
- Alternance entre apports théoriques et exercices pratiques (en moyenne sur 30 à 50% du temps)

Modalités pédagogiques : Présentiel, Distanciel et AFEST

Moyens techniques

En formation présentielle

Accueil des apprenants dans une salle dédiée à la formation et équipée avec :

- Ordinateurs
- Vidéo projecteur ou Écran TV interactif
- Tableau blanc ou Paper-Board

En formation distancielle

A l'aide d'un logiciel comme ® Microsoft Teams ou Zoom, un micro et une caméra pour l'apprenant.

- Suivez une formation en temps réel et entièrement à distance. Lors de la session en ligne, les apprenants interagissent et communiquent entre eux et avec le formateur.
- Les formations en distanciel sont organisées en Inter-Entreprise comme en Intra-Entreprise.
- L'accès à l'environnement d'apprentissage (support de cours, ressources formateur, fichiers d'exercices ...) ainsi qu'aux preuves de suivi et d'assiduité (émargement, évaluation) est assuré.
- Les participants recevront une convocation avec le lien de connexion à la session de formation.
- Pour toute question avant et pendant le parcours, une assistance technique et pédagogique est à disposition par téléphone au 02 35 12 25 55 ou par email à commercial@xxlformation.com

Modalités d'évaluation

- Positionnement préalable oral ou écrit.
- Feuille de présence signée en demi-journée.
- Evaluation des acquis tout au long de la formation.
- Questionnaire de satisfaction
- Attestation de stage à chaque apprenant
- Evaluation formative tout au long de la formation.
- Evaluation sommative faite par le formateur.

Profil du formateur

- Nos formateurs sont des experts dans leurs domaines d'intervention
- Leur expérience de terrain et leurs qualités pédagogiques constituent un gage de qualité

Adaptation pédagogique et matérielle

Si vous avez besoin d'adaptation matérielle ou pédagogique, merci de prendre contact avec notre référent Handicap par téléphone au 02 35 12 25 55 ou par email à handicap@xxlformation.com

Modalités et délais d'accès à la formation

Les formations sont disponibles selon les modalités proposées sur la page programme. Les inscriptions sont possibles jusqu'à 48 heures ouvrées avant le début de la formation. Dans le cas d'une formation financée par le CPF, ce délai est porté à 11 jours ouvrés.

Les conditions d'accès à la formation dans le cadre du CPF sont celles implicitement présentées dans les CGU disponibles sur le site Mon Compte Formation en vigueur à date d'inscription.

Voir CG_MCF_V16_VF.pdf Article 5.1.

Nos sessions INTER 2026

Sessions de formation à venir :

- Aucune session à venir pour cette formation.

Nos sessions INTRA 2026

Pour organiser cette formation en intra-entreprise, veuillez nous contacter par mail à commercial@xxlformation.com ou par téléphone au 02 35 12 25 55